

THE RELATIVE EFFICACY OF COMPARATIVELY WEAK AND STRONG
REINFORCEMENT CONTINGENCIES DURING SELF-EVALUATION
TRAINING

An abstract of a Thesis by
Robert F. Lynch
April 1981
Drake University
Advisor: W. Scott Wood

The problem. To assess the effectiveness of comparatively weak and strong reinforcement contingencies during self-evaluation training in producing maintenance of room cleaning behavior of youth in a shelter environment was the problem.

Procedure. Youth were trained to evaluate their own rooms and to determine the number of room points they would earn in a token economy based on that evaluation. Training procedures included, modeling, feedback following practice evaluation, and either token reinforcement for accurate evaluation or monetary reinforcement for accuracy. After training to criterion youth were allowed to assess their own room and reinforcement was based on this evaluation. The effect of varying the conditions under which training occurred was measured by monitoring performance when all of the obtrusive contingencies were withdrawn.

Findings. In three successive maintenance conditions dramatic reductions in the percentages of items cleaned according to the defined criteria occurred across seven of the eight subjects. Percentages of items cleaned appropriately were generally high for the group during all of the conditions in which token reinforcement was present. This included the two baseline/tokens conditions as well as the two self-evaluation conditions.

Conclusions. Token reinforcement in the form of points that could be exchanged for reinforcers consistently produced high cleaning percentages whether the condition existed alone or was combined with additional reinforcers for accurate self-evaluation. Self-evaluation did not significantly increase the youths cleaning behavior above the rates obtained during the baseline/tokens phase. Self-evaluation added desirable features to the token economy but did not contribute to any long term maintenance of the behavior when removed.

Recommendations. Researchers should continue to examine self-control strategies as potential techniques that may lead to generalization of treatment effects. Research studies may be most profitable if they experimentally verify the variables that operate when generalization is achieved.

THE RELATIVE EFFICACY OF COMPARATIVELY WEAK AND STRONG
REINFORCEMENT CONTINGENCIES DURING SELF-EVALUATION
TRAINING

A Thesis
Presented to
The School of Graduate Studies
Drake University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts

by
Robert F. Lynch
April 1981

THE RELATIVE EFFICACY OF COMPARATIVELY WEAK AND STRONG
REINFORCEMENT CONTINGENCIES DURING SELF-EVALUATION
TRAINING

by

Robert F. Lynch

Approved by Committee:

W. Scott Wood

Chairperson

Phillip K. Duncan

Lewis J. McMurran

Seale L. Canfield

Dean of the School of Graduate Studies

TABLE OF CONTENTS

	PAGE
INTRODUCTION AND REVIEW OF LITERATURE	1
METHOD	6
RESULTS	12
DISCUSSION	22
REFERENCES	29
APPENDIX	32

LIST OF TABLES

TABLE	PAGE
1. The mean and range of percentages of room areas cleaned for each subject across conditions	14
2. The percent agreement between youth and observer evaluations across self-evaluation conditions	15

LIST OF FIGURES

FIGURE	PAGE
1. The mean percent of room areas cleaned for all subjects as a function of Baseline/Tokens, Self-Evaluation, and Maintenance conditions	13

CHAPTER I

INTRODUCTION

A topic of continued concern to all behavior analysts has been maintenance and generalization of treatment effects. This same concern is an extremely important problem for practitioners and researchers involved in programming therapeutic environments such as token economies. In this paper maintenance will refer to the initial step in achieving stimulus generalization, the removal of a reinforcement contingency that has been shown to be functionally related to a change of behavior. Stimulus generalization will refer to the degree to which a change of behavior is likely to continue to occur as the stimulus complex that made up the treatment setting is varied along any number of dimensions (Skinner, 1953; Kazdin & Bootzin, 1972; Johnston, 1979).

Researchers and reviewers have pointed to the weakness of planning and programming for generalization following the removal of token reinforcement (Kazdin & Bootzin, 1972; O'Leary & O'Leary, 1976; Stokes & Baer, 1977). These articles describe such efforts to achieve generalization as programming for naturally occurring reinforcers, gradually withdrawing token control, and training significant others to maintain changes. In spite of these efforts, the authors note that active planning for generalization remains

absent and the process of stimulus generalization is often considered to be a natural or passive process.

To help achieve generalization these reviewers have advocated the implementation of "self-control" training. The self-control strategies that have been suggested are commonly described as self-assessment, self-recording, self-evaluation, and self-reinforcement (Kanfer & Duerfeldt, 1967; Bandura & Perloff, 1967; O'Leary & Drabman, 1971). It is difficult to separate any one of these strategies from the others in that they are often interdependent components of a comprehensive self-control training program. Nevertheless, there have been efforts to operationally classify these strategies (Bandura & Perloff, 1967; Glynn, Thomas & Shee, 1973; O'Leary & O'Leary, 1976; O'Leary & Dubey, 1979). These efforts may be premature at this point since the relative efficacy of the components has not been adequately demonstrated.

The large volume of literature in this area might lead some readers to conclude that the "self," apart from environmental variables, is in fact the container of variables to be manipulated. Further, they may allow for the implication that a "self" is, in fact, operative. A more appropriate behavioral perspective has been presented by Skinner (1953) and by Goldiamond (1976).

It has been suggested that the most salient behavior among the complex chain of self-controlling behaviors may

be accurate self-evaluation (Kanfer & Duerfeldt, 1967). This area of research typically includes the component behaviors of self-assessment and self-recording, where the subject's self-evaluation becomes the basis for determining the magnitude of reinforcement. Many research studies have investigated this strategy.

In the classroom, maintenance of behavior change during self-evaluation, following token economy interventions, has been reported by Bolstad and Johnson (1972), Kaufman and O'Leary (1972), Glynn, Thomas and Shee (1973) and by Drabman, Spitalnik and O'Leary (1973). These studies demonstrate behavior changes to be maintained when self-evaluation phases were introduced in combination with already existing reinforcement contingencies. However, in the studies just mentioned, decrements in classroom behavior and/or academic output often occurred when the baseline conditions were again implemented. Santogrossi, O'Leary, Romanczyk and Kaufman (1973) attempted but were unable to replicate the findings of Kaufman and O'Leary (1972).

Many researchers have attempted to assess the effects of self-evaluation alone. Turkewitz, O'Leary and Ironsmith (1975) found that self-evaluation was not effective in maintaining behavior change when it was introduced prior to any other intervention. The authors report that self-evaluation facilitated maintenance when it was introduced a second time following a reinforcement contingency that had produced

a behavior change. Both Fixsen, Phillips and Wolf (1972) and Layne, Richard, Jones and Lyman (1976) report that they were unable to achieve maintenance of room cleaning by youth in residential settings after the youth had been trained to accurately evaluate their rooms.

Apparently Wood and Flynn (1978) is the only study reporting long-term maintenance of behavior change as a function of self-evaluation training. In this study, maintenance of room cleaning behavior occurred for periods of up to 60 days. Youth who were described as "pre-delinquent" were first exposed to a token system that altered room cleaning behavior. The youth were then trained to accurately evaluate their own rooms with a procedure that shaped accuracy to a pre-set criterion. The self-evaluation component was then gradually faded out and maintenance of behavior changed continued.

Of particular interest in analyzing this study was the powerful contingency that was arranged during the self-evaluation phase. During this condition youth were allowed access to reinforcement periods in an all-or-none fashion depending upon the number of "clean room" points that had been earned on specified days. Room points were dependent both upon the cleanliness of the room as well as the accuracy of the self-evaluation. This study seems significant in that it appears to be the only study where maintenance of treatment effects have continued when baseline conditions

were reinstated. Other studies mentioned above have failed to achieve maintenance during a return to baseline, or simply did not reinstate baseline conditions.

An analysis of the functional variables in the Wood and Flynn (1978) study would increase our understanding of the conditions that are necessary and sufficient for maintenance following self-evaluation. One might hypothesize that an important variable in determining the likelihood of maintenance is the arrangement of reinforcement contingencies during the self-evaluation training phases of a study. It seems that the particularly strong contingency that was arranged by Wood and Flynn distinguishes this study from others that failed to achieve maintenance.

In summary, it appears that self-evaluation may lead to maintenance of appropriate behavior when: (1) change has been produced by a token program or some other contingency management system, (2) accurate self-evaluation skills are systematically shaped, and (3) the conditions necessary for shaping are gradually removed. The question regarding the necessary conditions for achieving these results with problem children remains. Therefore, the purpose of this study will be to compare relatively "weak" and "strong" reinforcement contingencies during self-evaluation phases, and to attempt to replicate the findings of Wood and Flynn (1978).

CHAPTER II

METHOD

Subjects

The subjects were selected from all the youth at the Polk County Juvenile Home Shelter Care facility. The subject population in this short term facility ranged between eight and seventeen years with a distribution of five boys and three girls in the study. The maximum census of twenty youth was attained frequently during the study. Only eight youth were used as subjects due to the changing nature of the population. All of the youth were court ordered to the shelter as an interim placement due to different circumstances that prevented them from continuing to reside at home. The children were adjudicated as children in need of assistance by the local Juvenile Court. The mean length of stay for youth in shelter was 35 days at the time of the study.

Conditions

Baseline. During baseline conditions, the youth in the study were in a token economy. The token economy was a standard behavior management package where points were earned across the day for many behaviors such as social interactions with staff and peers and following basic routines necessary for group living arrangements. Points were exchanged throughout the day for recreational activities,

treats, extra privileges such as staying up an additional hour, and for the purchase of craft items. Eligibility for many weekend activities was also based upon each youth earning a specified percent of the daily points for a specific number of days prior to the activity.

The youth were able to earn up to 1000 points each day with 30 of these points available for room cleaning behavior. During the baseline conditions, staff began the day by awakening the youth at 7:30 a.m. and providing an instruction to the youth indicating that it was time to begin completing the room cleaning tasks. The time designated for room cleaning and personal care was between 7:30 and 8:00 a.m. prior to breakfast and between 8:30 and 9:00 after breakfast.

Staff were instructed during baseline to provide not more than two prompts to any one youth to continue cleaning and not more than two instructions to any youth specifying how to clean a certain area or item according to the definitions. Youth were allowed to leave their rooms to take care of personal needs or to go to a designated recreation area any time after 8:35 a.m. after he/she stated to a staff person that their room was clean. Youth were allowed to return to a room upon request any time prior to 9:30 a.m. when the youth were summoned to line up for school. During cleaning periods beginning at 7:30 a.m. staff systematically provided cleaning materials, such as dust mops, cleanser,

sanitizer, toilet brushes, and clean trash bags, to all youth. At 9:30 A.M. two staff independently checked all rooms. Staff used a rating sheet and recorded each item for each youth with the symbols reflecting whether the item had met the defined criterion.

At the conclusion of the room checks, the staff multiplied the number of checks by two and awarded the youth with that number of points. For example, if a youth completed 8 of 15 items adequately, he/she received 16 of the available room points. Points on the card were rounded to the nearest five, so the youth would have received 15 points in this example.

Self Evaluation I. During this phase youth were given specific instructions on the use of a self-check recording sheet. This consisted of two fifteen minute sessions where a youth was given a presentation on the definitions of a clean room followed by an opportunity to rate a room item by and receive feedback from a trained observer concerning the accuracy of the rating. After the initial training all of the conditions present during baseline were introduced once again with the following additions and alterations.

The self evaluation phase was adapted from Wood and Flynn (1978). Each youth was supplied with a rating sheet and room definitions each day at 8:30 a.m. and instructed to rate his/her room. The sheet was turned in any time after 8:35 a.m. When the sheet was complete each youth was allowed

to go to a recreation room or to care for personal needs. Room checks were made at 9:30 a.m., as in baseline. The youth were able to earn two types of points related to the accuracy of the self-evaluation and cleanliness of the room (Wood & Flynn, 1978). A youth was able to earn two cleanliness points for each of the fifteen items that met the defined criterion and also earn bonus points for accuracy. A youth was able to earn two bonus points for each item rating that matched the item rating of the staff observer. The bonus points, as well as the clean room points, were presented to the youth at a mid-morning school break.

When a youth had rated his room in agreement with the staff observer on 80% of the items for two consecutive days the checks were reduced to one within the next two days. This accuracy checking system was adapted from Drabman et al. (1973), Turkewitz et al. (1975), and Wood and Flynn (1978). The youth were provided with individual feedback on the accuracy of the checks on each day that comparison checks were scheduled. The feedback consisted of an item by item comparison and discussion. If on the third comparison check the youth again agreed with the staff recorder on 80% of the items he/she was informed that his/her ratings would determine the number of room cleaning points he/she could earn from that point on. The bonus points were presented when the youth turned in the rating sheet with all of the items rated as either completed or incomplete. The youth were

also informed that at any time a spot check may occur and if agreement was below 80% the daily matching procedure would once again be implemented.

Self Evaluation II. During this phase all of the procedures in the self-evaluation phase were implemented with the addition of a monetary contingency. This contingency consisted of a set of instructions to all youth informing them that money for outings could be earned by cleaning rooms according to the definitions and accurately evaluating one's own behavior. The same three accuracy check procedures were implemented and youth were able to once again match the staff's evaluation and earn room cleaning points. During this condition the bonus points were again presented for accuracy, however, the youth were informed that they could earn 2 cents for each two accuracy points that were earned so that 30 cents could be earned each day for a total of \$1.10 per week.

Maintenance. During this phase all token reinforcement contingencies were removed. Youth were told upon wake-up to prepare for breakfast and school but were given no instructions or prompts to clean their rooms. Room checks continued in the manner that was described earlier.

Reliability. Twice each week an independent observer selected three rooms at random and checked the room according to the room cleaning definitions. The data sheet was then compared item by item with that of the primary

observers and reliability was computed according to the formula of agreements over agreements plus disagreements and multiplied by 100. An agreement was defined as a check for a particular item on both the primary and independent observers data sheets. A disagreement was defined as any combination of a check and a zero for a particular item.

Recording Behavior. Throughout all of the experimental conditions staff checked all youth rooms each day using a checklist. Staff were trained to an 80% agreement criterion using the fifteen item list of definitions for youth rooms. The staff data sheets were used in self-evaluation conditions I and II to determine accuracy of youth self-evaluations and were used as the data source to evaluate the efficacy of the procedures.

Experimental Design

The experimental design was a repeated time series replication design. The operating token system was considered as the baseline phase with a withdrawal of all token conditions as the maintenance conditions. The two self-evaluation conditions were introduced sequentially with maintenance phases introduced prior to and after the self-evaluation conditions. The major intent of the design was to provide a basis for making comparisons between the two self-evaluation conditions in terms of the maintenance of room cleaning behaviors.

CHAPTER III

RESULTS

Figure 1 represents the percent of the total possible room areas that were cleaned according to the definitions. The data that are presented in Figure 1 have been combined across all subjects in the study. Individual means and ranges for each of the subjects in each of the experimental conditions are displayed in Table 1. Percent, rather than absolute numbers, were reported because some youth had the opportunity to clean 14 areas in a room, while others had an opportunity to clean 15 areas in a room. The percent of agreement between youth self-evaluations and observer evaluations during both self-evaluation conditions are reported in Table 2.

There were a total of 11 subjects who were in the facility when data collection began. Of that original group, five subjects remained at the facility during the 44 days of the study. An additional 19 subjects entered the facility at various points in the study and were included as subjects during their period of residence. Six of the original group and all but three of the youth who entered the facility at different points during the study were released prior to the conclusion of the study. The data that are presented are from subjects who were participants throughout the study, as well as the three subjects who

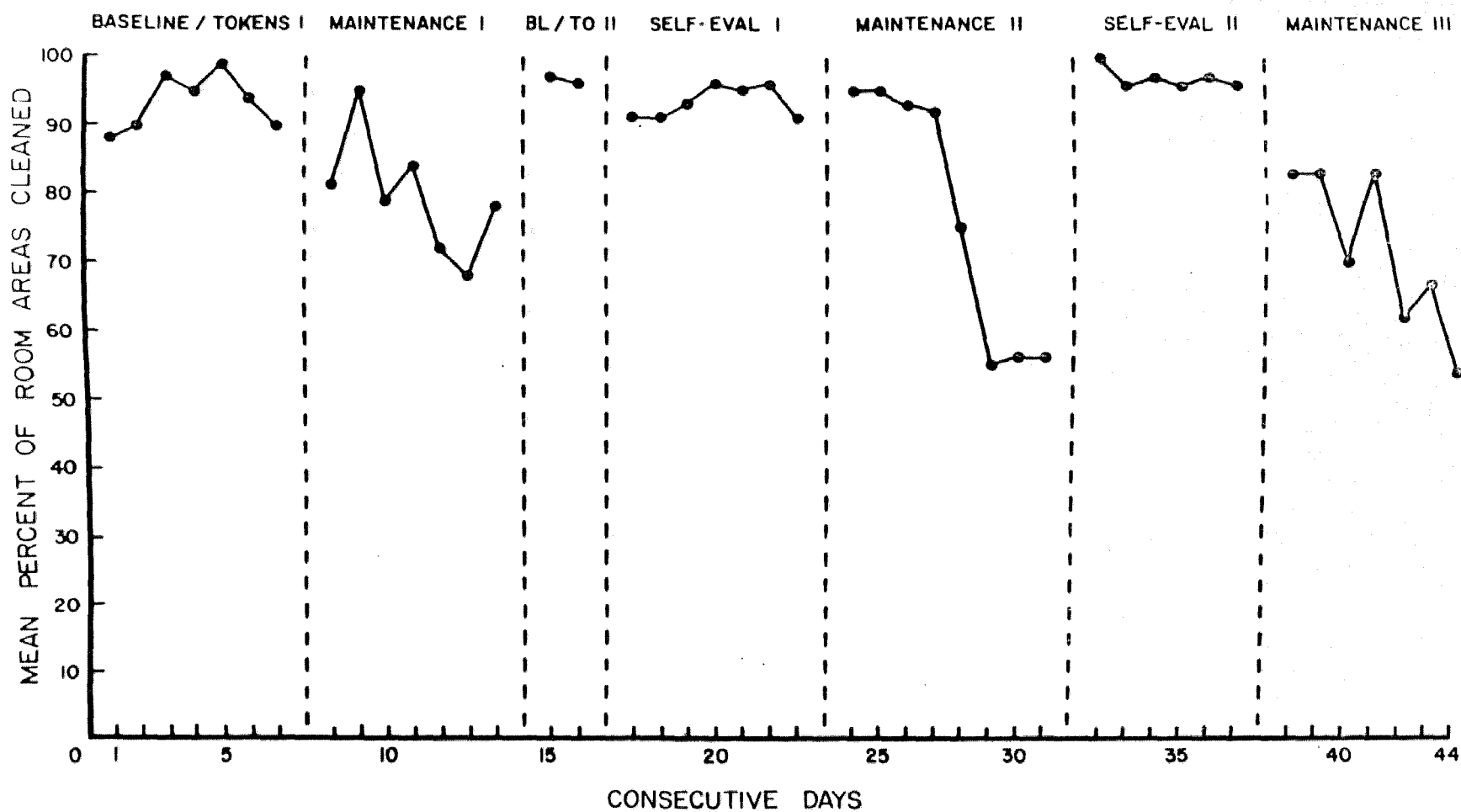


Figure 1. The mean percent of room areas cleaned for all subjects as a function of Baseline/Tokens, Self-Evaluation and Maintenance conditions.

Table 1

The Mean and Range of Percentages of Room Areas Cleaned for each
Subject Across Conditions

Youth	Conditions						
	BL/TO I	Main I	BL/TO II	S-E I	Main II	S-E II	Main III
Carol	$\bar{X} = 100$ 100-100	$\bar{X} = 92$ 86-100	$\bar{X} = 100$ 100-100	$\bar{X} = 100$ 100-100	$\bar{X} = 72$ 36-100	$\bar{X} = 99$ 93-100	$\bar{X} = 79$ 50-100
Mellisa	$\bar{X} = 97$ 86-100	$\bar{X} = 87$ 64-100	$\bar{X} = 100$ 100-100	$\bar{X} = 100$ 100-100	$\bar{X} = 80$ 50-100	$\bar{X} = 99$ 93-100	$\bar{X} = 74$ 57-100
Ben	$\bar{X} = 77$ 50-100	$\bar{X} = 51$ 47-79	$\bar{X} = 84$ 80-87	$\bar{X} = 76$ 64-87	$\bar{X} = 66$ 33-87	$\bar{X} = 87$ 87-87	$\bar{X} = 55$ 33-73
Arthur	$\bar{X} = 97$ 80-100	$\bar{X} = 90$ 80-100	$\bar{X} = 100$ 100-100	$\bar{X} = 99$ 93-100	$\bar{X} = 95$ 80-100	$\bar{X} = 97$ 93-100	$\bar{X} = 84$ 67-93
Kenny	$\bar{X} = 94$ 80-100	$\bar{X} = 67$ 53-100	$\bar{X} = 100$ 100-100	$\bar{X} = 79$ 87-100	$\bar{X} = 73$ 40-100	$\bar{X} = 98$ 93-100	$\bar{X} = 60$ 43-93
Cary	-	-	-	-	$\bar{X} = 55$ 40-80	$\bar{X} = 91$ 87-100	$\bar{X} = 69$ 47-87
Tracey	-	73	$\bar{X} = 94$ 87-100	$\bar{X} = 96$ 87-100	$\bar{X} = 89$ 73-100	$\bar{X} = 94$ 87-100	$\bar{X} = 85$ 67-100
Darrell	-	$\bar{X} = 63$ 53-67	$\bar{X} = 87$ 80-93	$\bar{X} = 69$ 67-87	$\bar{X} = 70$ 47-93	$\bar{X} = 82$ 60-93	$\bar{X} = 71$ 53-87

Table 2

The Percent Agreement Between Youth and Observer Evaluations
Across Self-Evaluation Conditions

Youth	Self-Evaluation I					Self-Evaluation II				
Carol	100	100	100	-	-	100	100	100	-	-
Mellisa	100	100	100	-	-	100	100	100	-	-
Ben	71	71	86	86	86	86	86	86	-	-
Arthur	100	100	100	93	-	100	100	100	-	-
Kenny	86	100	87	-	-	100	100	-	-	-
Cary	-	-	-	-	-	93	-	93	100	-
Tracey	100	100	86	-	-	100	87	100	-	-
Darrell	66	87	87	87	-	93	73	73	80	93

entered the facility at some point after the study had begun and who remained until the conclusion.

Data from the subjects in residence when the study began, but who were released prior to the study's conclusion, were not included since the primary comparisons that were the purpose of the study could not be made. This also applies to subjects who entered the study and were released prior to the conclusion. The three subjects that were included in the later conditions of the study were added into the data analysis because comparisons between the two self-evaluation conditions could be made. Data from these subjects were added into the group data. The addition of these subjects did, in retrospect, provide for some control of possible order effects that may have been masked had only subjects that were in the entire study been included in the data analysis.

The purpose of the design was to provide a basis for comparing the conditions under which self-evaluation was trained. The first condition in the sequence of interventions is referred to as Baseline/Tokens and is representative of the existing management program when the investigation began. A typical token economy had been in place for many months with the youth earning points for room cleaning and exchanging the points for conditioned reinforcers. Maintenance conditions were interspersed between the original token conditions and the two self-evaluation conditions.

Percentage Data

The first three conditions with the presentation and removal of contingent tokens clearly demonstrates the sensitivity of the dependent measure to contingency manipulation. Four of the five original subjects completed more than 80% of the room areas on each of the days in the first condition. A greater degree of variability was observed for Ben. This may be due to the fact that the boy had only recently been admitted and was only eight years old.

For all subjects the effect of the withdrawal of tokens and the reintroduction of the Baseline/Tokens condition demonstrates the responsiveness of the measure, and in so doing, the effectiveness of the token contingency in modifying room cleaning behavior. This effect is illustrated in the group data where the youth completed a mean of 93% of the items in Baseline/Tokens I, 80% in Maintenance I, and 98% of the items complete in Baseline/Tokens II.

The introduction of self-evaluation training in Self-Evaluation conditions I and II produced consistently high levels of performance for all subjects. The group data indicate that youth generally cleaned greater than 90% of the room areas in each of these conditions. This is representative of all subjects with the exception of Ben and Darrell where wider ranges within each condition were observed. It is interesting that for Ben, the wide range of percentages that occurred during both Baseline/Tokens I and

II and Self-Evaluation I is completely absent in Self-Evaluation II.

A similar pattern of increased and decreased performance that was described in the first three conditions, is repeated in the final four conditions. For all subjects, the introduction of Self-Evaluation I and Self-Evaluation II had the nearly identical effect of establishing reinforcement contingencies that produced high percentages of completion. Wider ranges in the percentage data for Ben and Darrell continue to be noticeable in Table 1. The efficacy of the training procedure in producing high percentages is indicated by the reduction in cleaning behaviors when Maintenance conditions II and III are introduced.

In order to compare the lasting effect of the Self-Evaluation conditions the critical comparisons are to be made between Maintenance conditions I, II, and III. The effect that the introduction of a Maintenance condition has on the room cleaning percentages is used to assess the efficacy of the preceding reinforcement condition in producing a desired long-term effect. During all of the Maintenance conditions, clear drops in percentages were observed for all subjects. The drops tended to be small but noticeable for subjects such as Arthur, where the lowest percentage in the three maintenance conditions was 80%, 80%, and 67%, respectively. Other subjects such as Kenny show much more dramatic declines, with the lowest percentage for

the three successive maintenance conditions falling to 53%, 40%, and 43%. The group data provide a clear representation of the declining percentages. The mean percentages for the three successive conditions was 80%, 77%, and 71%.

A brief maintenance effect during Maintenance II can be readily observed in the group data where the first four data points in the condition are above 90%, without any significant decreasing trend. Following this insignificant period in which the data points overlapped with all of the data points in the Self-Evaluation I condition, the data show that room cleaning percentages decreased to levels comparable to those observed in the other two maintenance conditions.

Maintenance of room cleaning did occur for one of the participants. The data in Table 1 for Tracey demonstrate that following the introduction of point reinforcement contingencies she continued to clean a high percentage of the room areas throughout the remainder of the study. The only data point that fell below 73% occurred during the final Maintenance condition when Tracey was informed that her release from the facility was imminent.

The data for Cary demonstrate that the same characteristic pattern of increasing decreasing percentages tended to occur even though the subject had not experienced the previous conditions of intervention and withdrawal. There appears to have been no advantage in terms of maintenance to

introduce only a single self-evaluation condition. These data confirm a conclusion that might have been drawn from all of the other subjects, namely, the order of the Self-Evaluation conditions was not a significant variable in producing maintenance. This conclusion would not have been possible without a subject receiving only one of the training conditions, or receiving the training in a different order.

Self-Evaluation Reliability

Table 2 presents the reliability coefficients that were obtained during both self-evaluation conditions. The coefficients were obtained by comparing the self-ratings made by youth with the data collection instrument that was used by staff throughout the study. Self-evaluation training required that each youth obtain an agreement coefficient of 80% or greater for two consecutive days before the youths' ratings would be used to determine the number of room cleaning points that were earned. In addition each youth was informed that a third unannounced check would be made and that 80% agreement was necessary if the youth was to continue as a self-evaluator.

The coefficients for each youth are presented. If the first two checks were greater than 80% then a third check was made within two days of the last check. If the first two checks were lower than 80%, reliability continued to be assessed on each succeeding day until the two consecutive

day criterion had been met. When this had occurred a third check was made within two days.

The reliability coefficients were generally quite high. Five of the eight youth met the criterion for self-evaluation within the shortest period possible during Self-Evaluation I, while seven of the eight youth met this same criterion during Self-Evaluation II. Of the 50 reliability coefficients that are reported, 25 are 100% agreement.

Reliability Data

The interobserver agreement between the daily staff observers and an independent observer was assessed 12 times during the study. Reliability data were recorded at least once during each of the experimental conditions. Of the 12 coefficients reported, only one fell below 80%, with agreement on 75% of the room areas. The remaining 11 coefficients ranged between 80% and 97% agreement.

CHAPTER IV

DISCUSSION

The purpose of this study was to compare the maintenance effect of comparatively weak and strong reinforcement contingencies during self-evaluation training and to attempt to replicate the findings of Wood and Flynn (1978). With respect to the comparison of training conditions, the data clearly show that the strength of the reinforcement contingency that was present during self-evaluation training was not a variable that added significantly to the maintenance of behavior when either the weak or strong token contingency was removed. In both of the conditions where self-evaluation was trained, room cleaning behaviors increased well above the previous maintenance condition, however similar patterns of decline were observed following both training conditions when maintenance conditions were once again introduced.

The study did not replicate the results of the research by Wood and Flynn (1978). In that study two groups of subjects maintained room cleaning behavior for periods of 22 and 60 days. One reason for the failure to replicate may have been the difference in the subject population. In the Wood and Flynn study subjects were described as "pre-delinquent," while the subjects in this study represented a more heterogeneous group. They had been ordered to the

shelter facility for many reasons that most often were not related to delinquent behavior. This difference in subject population may have been significant. Another difference was the long-term treatment facility used in the Wood and Flynn study. The facility in this study was a short-term shelter environment. One may suppose that youth in a facility where certain behavior changes are most likely necessary for release may be living under conditions where behavior change is more likely to be maintained.

In a broader conceptual context, O'Leary and Dubey (1979) have suggested three variables that may influence the efficacy of what they call self-assessment procedures. A review of these variables may help in understanding the differences in the results of this study and the Wood and Flynn (1978) study. The variables that were given primary focus by O'Leary and Dubey (1979) were: (1) the accuracy of the self-assessment, (2) the difficulty of the task or behavior, and (3) the type of child involved in the study. A comparison of this study and the Wood and Flynn results with respect to each of the variables provided by O'Leary and Dubey will follow.

Data in the results section of this study show the percent of agreement between the youths' self-evaluations and the independent staff observations. The data indicate that the youth tended to evaluate their rooms accurately. The reliability coefficients were frequently greater than

90%. Wood and Flynn (1978) report similar results with accuracy averaging 93% following evaluation training. The two studies appear to be nearly identical in terms of this variable.

O'Leary and Dubey (1979) cited a study by Peacock, Lyman and Richard (1978), to support their contention that maintenance becomes less likely as task difficulty increases. Obviously, this variable cannot account for any difference between the studies being compared. In fact, it seems reasonable to infer that the task was not difficult since the behaviors were usually a part of the subjects' repertoires prior to the study, or they became a part of their repertoires after the first few days of contingent reinforcement.

With respect to the third variable listed above, the type of child involved in the study, this study and the Wood and Flynn study differ in an important way. The difference in the subject population has been mentioned above. Again, the children in this study were residents in a short-term shelter compared with youth residing in a long-term treatment environment.

These comparisons would suggest that the two studies differed in only one way. A closer examination of the differences in the types of facilities where the youth lived during the respective studies perhaps suggests not only the critical variable in determining the efficacy of self-

evaluation, but also points up the inadequacy of the analysis provided by O'Leary and Dubey. It seems that by focusing on the type of child as one of three critical variables, O'Leary and Dubey have overlooked a more significant variable, namely, the conditions under which training and maintenance occur. O'Leary and Dubey describe how researchers have effectively implemented self-evaluation procedures after initially modifying behaviors through programmed reinforcement systems, that presumably changed their motivational levels and altered the "type" of child in the study. Considering that many studies have produced either very short periods of maintenance (e.g., Seymour & Stokes, 1976) or have not introduced withdrawal or extinction conditions (e.g., Kaufman & O'Leary, 1972), one should not be misled into thinking that once the "type" of child has been changed through token reinforcement that the mere introduction of self-evaluation will ensure maintenance of behavior changes. Given the many similarities between this study and the study by Wood and Flynn (1978) one must surely look upon the dramatic differences in the maintenance data as being a function of more permanent and stable aspects of the living environment. Clearly, the type of child involved in the study is a variable that should be considered far less significant than the conditions under which training and maintenance will occur.

It should also be noted in the context of similar

research in self-control that the results of this study are much more typical than the results of Wood and Flynn (1978). Aside from the Wood and Flynn study, only the study by Seymour and Stokes (1976) presents data showing behavior change to be maintained during periods of extinction or withdrawal of reinforcement contingencies. Even the data in the Seymour and Stokes study are not nearly as dramatic as the Wood and Flynn results since maintenance continued for periods of six or fewer days during "self-record only" conditions, and data were not recorded in the absence of all treatment interventions, including self-recording.

In the current study, a brief period of maintained room cleaning did occur during Maintenance II condition. This effect was of little practical value but was at least as significant as the maintenance data reported by Seymour and Stokes (1976). This statement is based on the fact that Seymour and Stokes did not introduce a true maintenance condition, as defined earlier, in that maintenance data were reported as occurring during a condition where subjects continued to self-record.

A comparison of the maintenance conditions in this study clearly shows that the addition of a weak or strong reinforcement contingency did not increase the probability of long-term behavior change. One clear conclusion that can be drawn from these data is that the performance of the youth was a direct function of the presence or absence of

programmed reinforcement.

While the data do not show much difference between the Baseline/Tokens conditions and the two Self-Evaluation conditions, some interesting and desirable effects were produced by self-evaluation training. The first effect was a greater degree of independent work by the youth. It had been typical of the morning routine for staff to frequently remind youth to clean their rooms. Problem behaviors were also commonly observed during this period. When youth were given the opportunity to rate their own rooms they began to work for longer and longer periods of time without prompting. This also resulted in a noticeable decrease in the frequency of disruptive behaviors.

With the increase in the amount of time that the youth spent working independently came a comparable decrease in the amount of time that staff were required to devote to monitoring youth behavior during this morning period. Because the staff spent less time dealing with disruptive behavior and prompting cleaning behavior, they were able to spend an increased amount of time interacting positively with the youth. The net effect of the self-evaluation training was to transform the most disruptive period of the day into one of the more orderly daily periods. The increased time that staff spent training self-evaluation was more than made up for during the more orderly morning period.

A general comment on the use of self-control training

can be made. Progress in behavior analysis has always followed from analyses that have been characterized by accounts of behavior that are stated in terms of orderly relations between observable behavior and environmental events (Skinner, 1953). Progress in the use of self-control strategies will most likely benefit from similar analyses. Goldiamond (1976) has argued that self-control strategies have great promise because "something else" operates when effective changes are produced. That "something else" is not presumed to reside within the organism but rather in a more complex set of environmental contingencies that lead to behavior change when an independent agent is not providing the evaluation or providing the reinforcement. At the same time Goldiamond cautions against any assumptions that self-control strategies are automatic panaceas for the problems of generalization and maintenance. The effective use of these behavior change strategies seems to hinge on the degree to which researchers and practitioners are able to provide complete behavior analyses of studies and cases where self-control strategies are employed.

REFERENCES

- Bandura, A., & Perloff, B. Relative efficacy of self monitored and externally imposed reinforcement systems. Journal of Personality and Social Psychology, 1967, 7, 111-116.
- Bolstad, O. D., & Johnson, S. M. Self regulation in the modification of disruptive classroom behavior. Journal of Applied Behavior Analysis, 1972, 5, 443-454.
- Drabman, R. S., Spitalnik, R., & O'Leary, K. D. Teaching self-control to disruptive children. Journal of Abnormal Psychology, 1973, 82, 10-16.
- Fixsen, D. L., Phillips, E. L., & Wolf, M. M. Achievement Place: The reliability of self-reporting and peer-reporting and their effects on behavior. Journal of Applied Behavior Analysis, 1972, 5, 19-30.
- Glynn, E. L., Thomas, J. D., & Shee, S. M. Behavioral self-control of on-task behavior in an elementary classroom. Journal of Applied Behavior Analysis, 1973, 6, 31-47.
- Goldiamond, I. Self-reinforcement. Journal of Applied Behavior Analysis, 1976, 9, 509-514.
- Johnston, J. M. On the relation between generalization and generality. The Behavior Analyst, 1979, 2, 1-6.
- Kanfer, F. H., & Deurfeldt, P. H. Effects of pretraining on self-evaluation and self-reinforcement. Journal of Personality and Social Psychology, 1967, 7, 164-168.

- Kaufman, K. F., & O'Leary, K. D. Reward, cost, and self-evaluation procedures for disruptive adolescents in a psychiatric hospital school. Journal of Applied Behavior Analysis, 1972, 5, 293-309.
- Kazdin, A. E., & Bootzin, R. R. The token economy: An evaluative review. Journal of Applied Behavior Analysis, 1972, 5, 343-372.
- Layne, C. C., Richard, H. C., Jones, M. T., & Lyman, R. D. Accuracy of self-monitoring on a variable ratio schedule of observer verification. Behavior Therapy, 1976, 7, 481-488.
- O'Leary, K. D., & Drabman, R. Token reinforcement programs in the classroom: A review. Psychological Bulletin, 1971, 75, 379-398.
- O'Leary, S. G., & Dubey, D. Applications of self-control procedures by children: A review. Journal of Applied Behavior Analysis, 1979, 12, 449-465.
- O'Leary, S. G., & O'Leary, K. D. Behavior modification in the school. In H. Leitenberg (Ed.), Handbook of behavior modification and behavior therapy. Englewood Cliffs, New Jersey: Prentice-Hall, 1976.
- Peacock, R., Lyman, R. D., & Richard, H. C. Correspondence between self-report and observer-report as a function of task difficulty. Behavior Therapy, 1978, 9, 578-583.
- Santogrossi, D. A., O'Leary, K. D., Romanczyk, R. G., & Kaufman, K. F. Self-evaluation by adolescents in a

psychiatric hospital school token program. Journal of Applied Behavior Analysis, 1973, 6, 277-287.

Seymour, F. W., & Stokes, T. F. Self-recording in training girls to increase work and evoke staff praise in an institution for offenders. Journal of Applied Behavior Analysis, 1976, 9, 41-54.

Stokes, T. F., & Baer, D. M. An implicit technology of generalization. Journal of Applied Behavior Analysis, 1977, 10, 349-367.

Skinner, B. F. Science and human behavior. New York: Macmillan, 1953.

Turkewitz, H., O'Leary, K. D., & Ironsmith, M. Generalization and maintenance of appropriate behavior through self-control. Journal of Consulting and Clinical Psychology, 1975, 43, 577-583.

Wood, R., & Flynn, J. M. A self-evaluation token system versus an external evaluation token system alone in a residential setting with predelinquent youths. Journal of Applied Behavior Analysis, 1978, 11, 503-512.

APPENDIX A

ROOM CLEANING DEFINITIONS

1. The bed is to be made with two sheets, a bed spread, a pillow case, and a blanket if desired by the youth.
There should be one pillow on the bed.
2. The bed is made with a fitted sheet over the mattress, a second sheet placed evenly over the mattress and tucked in on both sides and at the foot of the bed.
The bedspread should cover the sheets and should be tucked in so that only the spread is visible when an observer views the mattress at eye level. The pillow is to be centered at the head of the bed with approximately two inches of the spread tucked under the pillow and the remaining portion of the spread covering the pillow.
3. The bed should be made so that there are no wrinkles longer than three inches on the top surface of the bed.
4. There should be no objects on the bed other than specifically allowed stuffed animals.
5. Youth on Gold point card may have one additional set of clothing (top shirt, pants, socks, underwear, shorts) folded neatly in the youth's designated dresser drawer.
6. The dresser tops must be cleaned of all objects except pictures, models, craft items, ceramic figures.
7. Any other objects in the youth's room; e.g., writing paper, drawings, one pencil or pen, up to 3 books or 6 magazines, must be neatly stacked in the youth's dresser

or desk. Marking pens, model paints, model glue, scissors, and similar items are prohibited.

8. The floor is to be free of all debris including paper, personal clothing or shoes, or any other objects.
9. All room furniture is to be within six inches of but not touching the walls. Chairs should be pushed under desks where applicable.
10. The waste basket is to be empty with a clean plastic liner placed in it each morning during cleaning.
11. Posters or art objects are to be hung only in designated areas. Writing on or defacing walls is prohibited.
12. The bathroom fixtures; e.g., sink, toilet, mirror, are to be free of stains and streaks and soiled areas. Both youth within a room share the responsibility for cleaning these areas daily.
13. Each youth is to store his/her comb, soap, on the ledge attached to the mirror or the bathroom shelf. Each youth is to store his/her toothbrush and towel and cups in the toothbrush holder and the towel rack.
14. Curtains in the rooms are to be allowed to hang freely and pushed to the sides but not draped around bed posts or other available accessories.
15. Windows may be raised to designated levels and securely bolted. Storm windows may be raised or lowered, however, screens must remain lowered. Screens should be free of damage.